Was bedeutet p<0,05?

Oliver Kuß
Institut für Medizinische Epidemiologie, Biometrie und Informatik, Medizinische Fakultät,
Martin-Luther Universität Halle-Wittenberg, Halle (Saale)

Programm

- Einleitung
- Der Dopingtest
- Der statistische Test
- Das Problem
- Lösungen
- Fazit

Einleitung: Aus medizinischen Fachzeitschriften …

Special Article

What Your Statistician Never Told You about *P*-Values

Jeffrey Blume, Ph.D., and Jeffrey F. Peipert, M.D., MPH

J Am Assoc Gynecol Laparosc 2003;10:439-444

A Dirty Dozen: Twelve *P*-Value Misconceptions

Steven Goodman

Semin Hematol 2008;45:135-140

V: p-Werte: Was sie besagen und was nicht …

Frank Krummenauer

V: p Values: What they do Tell and what they Don't …

Klin Monbl Augenheilkd. 2002;219:896-8

Why *P* Values Are Not a Useful Measure of Evidence in Statistical Significance Testing

Raymond Hubbard
DRAKE UNIVERSITY

R. Murray Lindsay
UNIVERSITY OF LETHBRIDGE

Theory & Psychology 2008;18: 69-88.

What Is the Value of a *p* Value?

Gary L. Grunkemeier, PhD, YingXing Wu, MD, MS, and Anthony P. Furnary, MD
Medical Data Research Center, Providence Health & Services, Portland, Oregon

Ann Thorac Surg 2009;87:1337-43

The *p*-value – a well-understood and properly used statistical concept?

Oliver Kuss[1] and Andreas Stang[2]
[1] *Institut für Medizinische Epidemiologie, Biometrie und Informatik, Universitätsklinikum Halle und Medizinische Fakultät, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany and* [2] *Institut für Klinische Epidemiologie, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany*

Contact Dermatitis 2011;66:1-3

Einleitung: Aus dem statistischen Schrifttum…

TESTS OF SIGNIFICANCE CONSIDERED AS EVIDENCE*

BY JOSEPH BERKSON, M.D.
Division of Biometry and Medical Statistics, Mayo Clinic

J Am Stat Assoc 1942;37:325-335

Einleitung: Quiz

In a placebo-controlled trial of the use of aspirin and dipyridamole to prevent arterial restenosis after coronary angioplasty, 38% of patients receiving the treatment had restenosis, and 39% of patients receiving placebo had restenosis. In reporting this finding, the authors stated that $p>0.05$. This means

a. The chances are greater than 1 in 20 that a difference would be found again if the study were repeated.

b. The probability is less that 1 in 20 that a difference this large could occur by chance alone.

c. The probability is greater that 1 in 20 that a difference this large could occur by chance alone.

d. The chance is 95% that the study is correct.

Windish DM et al. JAMA. 2007;298(9):1010-1022

Einleitung: Quiz

Von 277 US-amerikanischen Assistenzärzten beantworten 59% [95%-KI: 53% - 65%] diese Frage richtig (Windish et al., 2007).

Von 208 französischen Intensivmedizinern beantworten 68% [95%-KI: 61% - 74%] diese Frage richtig (Darmon/Tabah, 2010).

Darmon M, Tabah A. Intens Care Med 2010; 35(Supp. 1):24

Einleitung: Quiz

Von 277 US-amerikanischen Assistenzärzten beantworten 59% [95%-KI: 53% - 65%] diese Frage richtig (Windish et al., 2007).

Von 208 französischen Intensivmedizinern beantworten 68% [95%-KI: 61% - 74%] diese Frage richtig (Darmon/Tabah, 2010).

Bittere Ironie: Die von Windish et al. vorgesehene Antwort ist falsch (Goodman, 2008)!

Darmon M, Tabah A. Intens Care Med 2010; 35(Supp. 1):24.

Einleitung

Bedingte Wahrscheinlichkeiten:

$P(A|B)$: Wahrscheinlichkeit für das Eintreten des Ereignisses A, wenn wir wissen, dass B wahr ist. („Wahrscheinlichkeit für A gegeben B")

Beispiel: Verdecktes Ziehen einer Karte aus einem Skatblatt (32 Karten)

Ereignis A: Karte ist ein Herz $\rightarrow P(A) = 8/32 = ??$

Ereignis B: Karte ist ein As $\rightarrow P(B) = 4/32 = ??$

$P(A|B)$: Wahrscheinlichkeit für ein Herz, wenn wir wissen, dass die Karte ein As ist $\rightarrow P(A|B) = ??$

$P(B|A)$: Wahrscheinlichkeit für ein As, wenn wir wissen, dass die Karte ein Herz ist $\rightarrow P(B|A) = ??$

Einleitung

Bedingte Wahrscheinlichkeiten:

$P(A|B)$: Wahrscheinlichkeit für das Eintreten des Ereignisses A, wenn wir wissen, dass B wahr ist. („Wahrscheinlichkeit für A gegeben B“)

Beispiel: Verdecktes Ziehen einer Karte aus einem Skatblatt (32 Karten)

Ereignis A: Karte ist ein Herz $\rightarrow P(A) = 8/32 = 1/4$

Ereignis B: Karte ist ein As $\rightarrow P(B) = 4/32 = ??$

$P(A|B)$: Wahrscheinlichkeit für ein Herz, wenn wir wissen, dass die Karte ein As ist $\rightarrow P(A|B) = ??$

$P(B|A)$: Wahrscheinlichkeit für ein As, wenn wir wissen, dass die Karte ein Herz ist $\rightarrow P(B|A) = ??$

Einleitung

Bedingte Wahrscheinlichkeiten:

$P(A|B)$: Wahrscheinlichkeit für das Eintreten des Ereignisses A, wenn wir wissen, dass B wahr ist. („Wahrscheinlichkeit für A gegeben B“)

Beispiel: Verdecktes Ziehen einer Karte aus einem Skatblatt (32 Karten)

Ereignis A: Karte ist ein Herz $\rightarrow P(A) = 8/32 = 1/4$

Ereignis B: Karte ist ein As $\rightarrow P(B) = 4/32 = 1/8$

$P(A|B)$: Wahrscheinlichkeit für ein Herz, wenn wir wissen, dass die Karte ein As ist $\rightarrow P(A|B) = ??$

$P(B|A)$: Wahrscheinlichkeit für ein As, wenn wir wissen, dass die Karte ein Herz ist $\rightarrow P(B|A) = ??$

Einleitung

Bedingte Wahrscheinlichkeiten:

$P(A|B)$: Wahrscheinlichkeit für das Eintreten des Ereignisses A, wenn wir wissen, dass B wahr ist. („Wahrscheinlichkeit für A gegeben B")

Beispiel: Verdecktes Ziehen einer Karte aus einem Skatblatt (32 Karten)

Ereignis A: Karte ist ein Herz $\rightarrow P(A) = 8/32 = 1/4$

Ereignis B: Karte ist ein As $\rightarrow P(B) = 4/32 = 1/8$

$P(A|B)$: Wahrscheinlichkeit für ein Herz, wenn wir wissen, dass die Karte ein As ist $\rightarrow P(A|B) = 1/4$

$P(B|A)$: Wahrscheinlichkeit für ein As, wenn wir wissen, dass die Karte ein Herz ist $\rightarrow P(B|A) = ??$

Einleitung

Bedingte Wahrscheinlichkeiten:

$P(A|B)$: Wahrscheinlichkeit für das Eintreten des Ereignisses A, wenn wir wissen, dass B wahr ist. („Wahrscheinlichkeit für A gegeben B“)

Beispiel: Verdecktes Ziehen einer Karte aus einem Skatblatt (32 Karten)

Ereignis A: Karte ist ein Herz $\rightarrow P(A) = 8/32 = 1/4$

Ereignis B: Karte ist ein As $\rightarrow P(B) = 4/32 = 1/8$

$P(A|B)$: Wahrscheinlichkeit für ein Herz, wenn wir wissen, dass die Karte ein As ist $\rightarrow P(A|B) = 1/4$

$P(B|A)$: Wahrscheinlichkeit für ein As, wenn wir wissen, dass die Karte ein Herz ist $\rightarrow P(B|A) = 1/8$

Der diagnostische Test

		Krankheit	
		K+	K-
Diagnostischer Test	T+	RP	FP
	T-	FN	RN

Maßzahlen für die Güte eines diagnostischen Tests:

Sensitivität: $P(T+ | K+)$

W't für positives Testergebnis bei einem Kranken

Spezifität: $P(T- | K-)$

W't für negatives Testergebnis bei einem Gesunden

Der Dopingtest

		Sportler gedopt	
		D+	D-
Dopingtest	T+	RP	FP
	T-	FN	RN

Maßzahlen für die Güte eines diagnostischen Tests:

Sensitivität: $P(T+ \mid D+)$

W't für positives Testergebnis bei einem Gedopten

Spezifität: $P(T- \mid D-)$

W't für negatives Testergebnis bei einem nicht Gedopten

Der Dopingtest

Aber:

Diese Werte nützen uns im Dopingsystem nichts, denn da haben wir nur das Testergebnis (T+ bzw. T-), wollen aber wissen, ob Doping vorliegt (D+ bzw. D-)

Von Interesse also: P(D+ | T+) bzw. P(D- | T-)

(Prädiktive Werte)

Der Dopingtest

Satz von Bayes:

$$P(D+ \mid T+) = \frac{P(T+ \mid D+)\, P(D+)}{P(T+ \mid D+)\, P(D+) + P(T+ \mid D-)\, P(D-)}$$

$$= \frac{\text{Sens x Prävalenz}}{\text{Sens x Prävalenz + (1-Spez) x (1-Prävalenz)}}$$

Problem: Prädiktive Werte sind prävalenzabhängig!

Der Dopingtest

Beispiel (ausgedacht):

Sensitivität:	$P(T+ \mid D+)$	= 95%
Spezifität:	$P(T- \mid D-)$	= 95%
Prävalenz:	$P(D+)$	= 10%

Wahrscheinlichkeit, dass jemand mit positivem Dopingtest gedopt ist:

$$P(D+ \mid T+) = \frac{P(T+ \mid D+)\, P(D+)}{P(T+ \mid D+)\, P(D+) + P(T+ \mid D-)\, P(D-)} = \frac{0{,}95 \times 0{,}1}{0{,}95 \times 0{,}1 + 0{,}05 \times 0{,}9} = 67{,}8\%$$

Von 3 positiv Getesteten sind nur 2 in Wahrheit gedopt (1 wird ungerechtfertigterweise gesperrt …)

Der statistische Test

		Wahrheit (Nullhypothese)	
		+ (H_0 richtig)	- (H_0 falsch)
Statistischer Test	ST+ (H0 akzeptieren)	☺	Fehler 2. Art
	ST- (H0 verwerfen)	Fehler 1. Art	☺

Fehler 1. Art: $\alpha = P(ST- \mid H_0+)$

W't für Verwerfen der Nullhypothese, wenn Sie vorliegt.

Fehler 2. Art: $\beta = P(ST+ \mid H_0-)$

W't für Annahme der Nullhypothese, wenn Sie nicht vorliegt.

Der statistische Test

Durchführung eines statistischen Tests:

- Lege α vor Beginn des Experimentes fest (z.B: $\alpha = 0.05$)
- Nach der Durchführung des Experiments (Daten: x) berechnen wir eine Teststatistik T(x) und entscheiden:

$$\text{Verwerfe } H_0\text{, falls } P(T >= T(x) \mid H_0+) < \alpha$$

$P(T >= T(x) \mid H_0+)$ ist der **p-Wert**.

Der p-Wert ist die Wahrscheinlichkeit, dass, unter der Annahme dass die Nullhypothese richtig ist, die Teststatistik einen Wert annimmt, der so extrem oder noch extremer als der tatsächlich beobachtete Wert ist.

Das Problem

Der p-Wert macht eine Aussage über die Wahrscheinlichkeit der Daten, nicht über die Wahrscheinlichkeit der Hypothese!!!

Wir möchten P(H0 | Daten), bekommen aber P(Daten | H0)!!!

Lösung (Anwendung des Satz von Bayes):

$$P(H_0+ \mid \text{Daten}) \approx \frac{P(\text{Daten} \mid H_0+)\, P(H_0+)}{P(\text{Daten} \mid H_0+)\, P(H_0+) + P(\text{Daten} \mid H_0-)\, P(H_0-)}$$

Großes Problem: Wie groß ist P(H0+)?
(A priori - Wahrscheinlichkeit für die Nullhypothese)

Lösungen

Beispiel: Randomisierte Studie mit zwei Behandlungen (Intervention/Kontrolle) und binärer Zielgröße (Erfolg ja/nein)

	Erfolg: Ja	Erfolg: Nein	Σ
Intervention	58	42	100
Kontrolle	42	58	100
Σ	100	100	200

Ergebnis: OR [95%-KI]: 1,91 [1,09; 3,34]

$\chi^2 = 5{,}12 \Rightarrow$ p-Wert = 0,024

Lösungen

Mögliche Lösung:

Nullhypothese und Alternative sind gleichwahrscheinlich ($P(H_0+) = 0{,}5$)

Für unsere Beispielstudie gilt dann: $P(H_0+ \mid \text{Daten}) = 0{,}135$

Spannende Frage:

Wie groß müsste $P(H_0+)$ sein, damit der p-Wert gleich $P(H_0+ \mid \text{Daten})$ wäre?

Antwort (für unser Beispiel)**:** $P(H_0+) = 13\%$

Berechtigte Frage: Ist es überhaupt ethisch, eine Studie zu machen, wenn man sich zu 87% sicher ist, dass die Intervention einen Effekt hat?

Lösungen

Weitere Lösungen (ohne Angabe von $P(H_0+)$!)

- **Minimum Bayes Factor** (Goodman, 1999):

$$P(H_0+ \mid \text{Daten}) = \frac{\exp(-\chi^2/2))}{(1+\exp(-\chi^2/2))}$$

In unserem Beispiel: $P(H_0+ \mid \text{Daten}) = 0{,}071$

- **Calibration of p-values** (Sellke/Bayarri/Berger, 1999):

$$P(H0+ \mid \text{Daten}) = (1-\exp(p\text{-Wert}*\log(p\text{-Wert}))^{-1})^{-1} \quad (\text{für } p\text{-Wert} < 1/e \ (= 0.368))$$

In unserem Beispiel: $P(H_0+ \mid \text{Daten}) = 0{,}194$

Lösungen: Auf die harte Tour …

„Significance ban“ durch den neuen Associate Editor (K. Rothman) des American Journal of Public Health.

Antworten an Autoren (Fleiss JL, 1986):
„All references to statistical hypothesis testing and statistical significance should be removed from the paper.“
„I ask you to delete *p* values as well as comments about statistical significance.“
„If you do not agree with my standards (concerning the appropriateness of significance tests), you should feel free to argue the point, or simply ignore what you may consider to be my misguided view, by publishing elsewhere.“

Fleiss JL. Am J Public Health 1986;76(5):559-60.

Lösungen: Auf die harte Tour …

Später: Kenneth Rothman als Editor von „Epidemiology".

When writing for **Epidemiology**, you can also enhance your prospects if you omit tests of statistical significance. Despite a widespread belief that many journals require significance tests for publication, the Uniform Requirements for Manuscripts Submitted to Biomedical Journals[3] discourages them, and every worthwhile journal will accept papers that omit them entirely. In **Epidemiology**, we do not publish them at all.

Rothman KJ. Epidemiology 1998;9:333-7.

Lösungen: Kompletter Umstieg in die Bayesianische Statistik

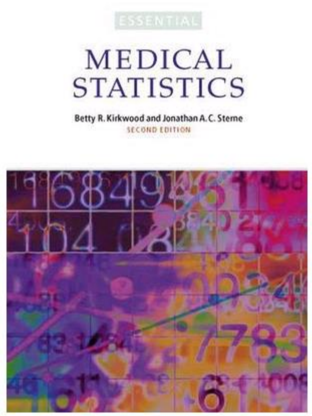
A priori-Wahrscheinlichkeiten ($P(H_0+)$) müssen vorher aufgrund von Vorwissen (aus Vorstudien, anderen Studien oder „informed guesses“) spezifiziert werden.

→ Wahrscheinlichkeitssaussagen über H_0 sind dann möglich.

Es können nicht nur A priori-Wahrscheinlichkeiten angegeben werden, sondern ganze A priori-Verteilungen, z.B. für Effekte.

Lösungen: Kompletter Umstieg in die Bayesianische Statistik

Ein Anfang ist gemacht:

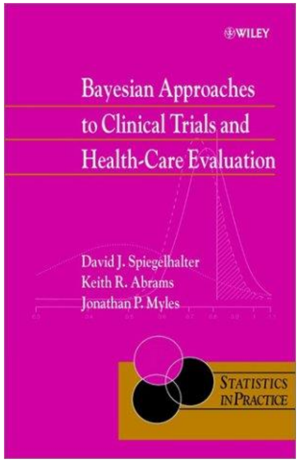


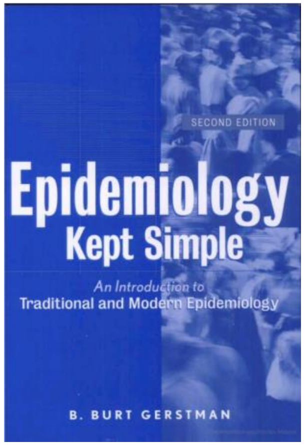
Aus dem Inhaltsverzeichnis:

.....

⟵ 33. Bayesian statistics (p. 388-391)

....

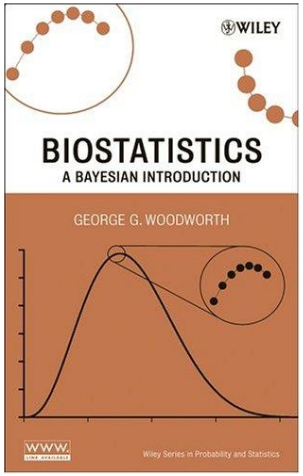




.....

⟵ 13.4. Minimum Bayes factors (p. 257-259)

....



Fazit

- **Der p-Wert sagt uns nicht das, was wir eigentlich wissen wollen, denn er macht keine Aussage über die Sicherheit der zugrunde liegenden Hypothesen.**
- **Die Sicherheit, die uns der p-Wert vorgaukelt, ist im allgemeinen zu groß.**

Schritte zu einer Lösung(?):

- Gesunde Skepsis gegenüber p-Werten
- Konfidenzintervalle berichten
- Für die Zukunft: Kompletter Umstieg in die Bayes'sche Statistik.

Auflösung Quiz

In a placebo-controlled trial of the use of aspirin and dipyridamole to prevent arterial restenosis after coronary angioplasty, 38% of patients receiving the treatment had restenosis, and 39% of patients receiving placebo had restenosis. In reporting this finding, the authors stated that p>0.05. This means

a. The chances are greater than 1 in 20 that a difference would be found again if the study were repeated.

b. The probability is less that 1 in 20 that a difference this large could occur by chance alone.

c. The probability is greater that 1 in 20 that a difference this large could occur by chance alone.

d. The chance is 95% that the study is correct.

Auflösung Quiz

In a placebo-controlled trial of the use of aspirin and dipyridamole to prevent arterial restenosis after coronary angioplasty, 38% of patients receiving the treatment had restenosis, and 39% of patients receiving placebo had restenosis. In reporting this finding, the authors stated that $p>0.05$. This means

a. The chances are greater than 1 in 20 that a difference would be found again if the study were repeated.

b. The probability is less that 1 in 20 that a difference this large could occur by chance alone.

c. The probability is greater that 1 in 20 that a difference this large could occur by chance alone.

d. The chance is 95% that the study is correct.

Auflösung Quiz

In a placebo-controlled trial of the use of aspirin and dipyridamole to prevent arterial restenosis after coronary angioplasty, 38% of patients receiving the treatment had restenosis, and 39% of patients receiving placebo had restenosis. In reporting this finding, the authors stated that $p>0.05$. This means

a. The chances are greater than 1 in 20 that a difference would be found again if the study were repeated.

b. The probability is less that 1 in 20 that a difference this large could occur by chance alone.

c. The probability is greater that 1 in 20 that a difference this large **or larger** could occur by chance alone.

d. The chance is 95% that the study is correct.